U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
PB80-132830

Spatial Variability of Ozone and Other Pollutants at St. Louis, Missouri

(U.S.) Environmental Sciences Research Lab, Research Triangle Pa
Research Triangle Park, NC

Oct 79

United States
Environmental Protection
Agency

Environmental Sciences Research
Laboratory
Research Triangle Park NC 27711

EPA-600/4-79-063
October 1979

Research and Development

PB80-132830

EPA

Spatial Variability of Ozone and Other Pollutants at St. Louis, Missouri

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations. It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/4-79-063		PB80-132830

4. TITLE AND SUBTITLE	5. REPORT DATE
SPATIAL VARIABILITY OF OZONE AND OTHER POLLUTANTS AT ST. LOUIS, MISSOURI	October 1979
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Thomas R. Karl	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
(Same as Block 12.)	1AA603 AE-11 (FY-79)
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Environmental Sciences Research Laboratory - RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC 27711	Inhouse 11/78-6/79
	14. SPONSORING AGENCY CODE
	EPA/600/9

15. SUPPLEMENTARY NOTES

16. ABSTRACT

A network of 25 aerometric stations was part of the U.S. Environmental Protection Agency's Regional Air Pollution Study (RAPS) in the greater St. Louis area. At these stations ozone (O_3) and various other photochemical pollutants (NO, NO_2, and total hydrocarbons (THC)) as well as carbon monoxide (CO) were analyzed with respect to their spatial variability. Data were analyzed for the warm months of the year, April through October of 1975 and 1976--periods during which high O_3 concentrations are common. The results of these analyses indicate that when O_3 concentrations are high (above 100 ppb) the daily 1-h maximum O_3 concentration is highly dependent upon the location of measurement. Measurements made at these times in both urban and rural locations in and around St. Louis were frequently found to vary by a factor of two and occasionally by a factor as large as four. Simultaneous hourly average measurements for each of the pollutants were correlated across the network of 25 stations. Rather poor correlations were found for the primary pollutants such as CO, NO, NO_2, and THC. Analyses of normalized fields of pollutant concentrations suggest that the scavenging of O_3 by NO dominates the O_3-NO_x reaction cycle before 1000 CDT.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
* Air pollution * Ozone Nitrogen oxides Hydrocarbons * Spatial distribution	St. Louis, MO	13B 07B 07C

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	42
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE PC-A03/MEA01

EPA Form 2220-1 (9-73)

EPA-600/4-79-063
October 1979

SPATIAL VARIABILITY OF OZONE AND OTHER POLLUTANTS AT ST. LOUIS, MISSOURI

by

Thomas R. Karl
Meteorology and Assessment Division
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

ENVIRONMENTAL SCIENCES RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

i-R

DISCLAIMER

This report has been reviewed by the Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Mr. Karl is a meteorologist in the Meteorology and Assessment Division, Environmental Sciences Research Laboratory, Environmental Research Center, Research Triangle Park, N.C. 27711. He is on assignment from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

ABSTRACT

A network of 25 aerometric stations was part of the U.S. Environmental Protection Agency's Regional Air Pollution Study (RAPS) in the Greater St. Louis area. At these stations ozone (O_3) and various other photochemical pollutants (NO, NO_2, and total hydrocarbons (THC)) as well as carbon monoxide (CO) were analyzed with respect to their spatial variability. Data were analyzed for the warm months of the year, April through October of 1975 and 1976--periods during which high O_3 concentrations are common. The results of these analyses indicate that when O_3 concentrations are high (above 100 ppb) the daily 1-h maximum O_3 concentration is highly dependent upon the location of measurement. Measurements made at these times in both urban and rural locations in and around St. Louis were frequently found to vary by a factor of two and occasionally by a factor as large as four. Simultaneous hourly average measurements for each of the pollutants were correlated across the network of 25 stations. Rather poor correlations were found for the primary pollutants such as CO, NO, NO_2, and THC. Analyses of normalized fields of pollutant concentrations suggest that the scavenging of O_3 by NO dominates the O_3-NO_x reaction cycle before 1000 CDT.

CONTENTS

Abstract. iii

Figures vi

Tables. vii

1. Introduction. 1

2. Conclusions 2

3. Data and Station Sites. 3

4. Procedures and Results. 4

Some elementary statistics on pollutant concentrations . 4

Maximum O_3 concentrations and the air flow 6

Hourly pollutant concentrations and the urban plume. . . 7

Cluster analysis of pollutant concentrations 9

5. Summary 14

References. 15

FIGURES

Number		Page
1	Stations of the Regional Air Monitoring Systems.	17
2	Annual total emissions for the St. Louis area. Diagrams on the right show emissions for 5-km grid squares in the more intense emission areas, while diagrams on the left are for 10- and 15-km grids covering the entire region. Emissions are directly proportional to the area covered by each asterisk; emissions are specified for the largest asterisk in each of the six diagrams.	18
3	One-H average O_3 concentrations and the highest 1-h O_3 concentration (ppb) between 1300 and 1400 CDT for the study period.	19
4	Hourly average concentrations of THC (ppm x 10^{-1}) CO (ppm x 10^{-1}) NO (ppb), and NO_2 (ppb) during 0700-0800 CDT.	20
5	Percentage of the highest O_3 concentration in the RAMS for individual stations on days of downwind and upwind flow . . .	21
6	Percentage of the highest O_3 concentration in the RAMS for individual stations on days of high wind speeds (>3.4 m/s) and on days of low wind speeds (<2.5 m/s).	22
7a.-7f.	Normalized concentrations for various pollutants. O_3, NO, and NO_2 concentrations are in ppb. THC and CO concentrations are in ppm x 10^{-1}. The zero isopleth is indicated on each map.	23
8a.-8f.	Diagram for cluster analysis of hourly pollutant concentrations and maximum O_3 concentrations. The first correlation coefficient inside each block represents the similarity of the two clusters which are combined into a single cluster. The second correlation coefficient (or first if the second is absent) is the minimum correlation coefficient between any two stations of the previously distinct clusters. The hour for which each cluster analysis is presented is that for which the network average concentration was highest.	29

TABLES

Number		Page
1	Percent of Daily Maximum 1-H O_3 Concentrations at Each Station That are Within Given Percentage Intervals of the Corresponding Daily Highest 1-H O_3 Concentration in the Network. N is the Number of Days of Valid Data for the Stations When the 1-H O_3 Concentration at Any RAMS Station Exceeded 100 PPB. .	32
2	Combinations of Stations Whose Concentrations are Most Frequently Within 80-100% of the Highest Daily Maximum 1-H O_3 Concentration in the Network (For Days With Maximum Concentration Above 100 PPB). N is the Total of Days Above 100 PPB and n is the Number of Days Which Were Within 80-100% of These High Concentrations at the Indicated Stations.	33

SECTION 1

INTRODUCTION

Attainment of the National Ambient Air Quality Standard (NAAQS) for ozone (O_3) is determined from measurements at fixed monitoring stations. A number of reports have either directly or indirectly confronted the problem of attaining the NAAQS for O_3 with missing or invalid measurements at stationary sites[1,2,3]. A somewhat similar, but more difficult problem concerns the representativeness of point measurements. This can also be regarded as a missing data problem, pertaining to space instead of time.

The spatial variability of O_3 and other pollutants affect the interpretation of output from photochemical and other air quality models. A question is raised regarding the representativeness of pollutant concentrations spatially averaged over some grid scale with respect to the actual concentrations observed in the atmosphere. Two other questions then arise. How much do stationary point measurements vary in space in and around metropolitan areas, and if mobile measurements cannot routinely be made, where and how many fixed point measurements are necessary to designate the area of highest O_3 concentrations? Answers to these questions were recently pursued also by others[4], using more limited data but with similar conclusions.

This study investigates the spatial variability of pollutant concentrations, O_3 in particular, using data from the Regional Air Monitoring System (RAMS) of the U.S. Environmental Protection Agency's Regional Air Pollution Study (RAPS) in the St. Louis, Missouri area. Using these data some answers are provided to the questions posed here.

SECTION 2

CONCLUSIONS

Primary pollutants, such as THC, CO, NO, and NO_2, require a dense network of stations in order to describe ambient concentrations across a metropolitan area such as St. Louis. Assuming St. Louis is fairly typical of many large cities, a much smaller number of carefully selected sites than the 25 employed in the RAPS would adequately monitor the maximum O_3 concentrations in and around metropolitan areas. The site selection process must include climatological information on the local air flow during meteorological regimes that are favorable for high O_3 concentrations. This information should be used in conjunction with emission data to locate the stations. The results demonstrate that to measure the regional maximum concentration of O_3, it would be misleading to rely on O_3 monitors located only in and around high emission areas.

From this study it is concluded that the grid-point output from photochemical models should have sufficient resolution to resolve the differences in O_3 concentration between areas of both concentrated and diffuse emissions or areas within and outside the urban plume. Otherwise, there is little chance of replicating area-wide maximum O_3 concentrations.

SECTION 3

DATA AND STATION SITES

The RAMS was a network of 25 stations in the St. Louis area (Figure 1). Wind speed, wind direction, O_3, carbon monoxide (CO), nitrogen dioxide (NO_2), nitric oxide (NO), and total hydrocarbons (THC) were continuously measured at each station. Wind speed and direction were measured atop 10-m or 30-m towers, while the effective measurement height for the pollutants was 4 m[5]. During the placement of these sites, domination of a site by one or more local sources was avoided[6].

One-minute averages of the data were automatically recorded on magnetic tape at RAPS headquarters. These averages were checked by various computer programs for erroneous data[7]. The data were screened for continuity, relational consistency, and the operational mode of the instruments. These data for the months of April through October of 1975 and 1976 were used to calculate hourly average quantities. A careful visual inspection of the hourly averages helped assure against the inclusion of fictitious data.

An extensive emission inventory for the Greater St. Louis area was compiled as part of the RAPS[8]. Figure 2 illustrates the spatial variability of total (area and point sources) annual emissions of THC, NO_x, and CO in the St. Louis area. This diagram is based on the annual emissions for 1975; however, there was no detectable difference when a similar diagram was constructed using 1976 emissions. Variable grid spacing was used to conduct the actual RAPS emission inventory. Grid squares were smallest (1 km) in areas with concentrated emissions and largest (10 km) in areas with less concentrated emissions. In order to visually display emissions from these variable grid squares they were apportioned out over a more uniform grid network. Grid squares of sizes 5, 10, and 15 km are contained in Figure 2. Apportionments were also made for 2.5 grid squares, but little additional detail could be found regarding the spatial variability of the emissions.

SECTION 4

PROCEDURES AND RESULTS

SOME ELEMENTARY STATISTICS ON POLLUTANT CONCENTRATIONS

Stations were checked each day for hourly average O_3 concentrations in excess of 100 ppb. When any of the 25 stations had concentrations this high the maximum O_3 concentrations for all the stations for the given day were set aside for further study. A concentration of 100 ppb was considered high because at the time of this research a new NAAQS for O_3 was proposed at such a level. A total of 154 days had at least one station exceeding an O_3 concentration of 100 ppb. The daily highest 1-h O_3 concentration at each station was converted to a percentage of the network highest 1-h O_3 concentration at any station for each of the 154 days, (this value will be referred to as O_3MAXP). For example, if station 107 had a 1-h O_3 concentration of 150 ppb on July 14, which was the highest 1-h O_3 concentration in the RAMS on that day, 100 would be the percentage of the highest 1-h O_3 concentration for station 107 on July 14. All other stations would have a smaller percentage associated with them for July 14, unless they had missing hourly average O_3 concentrations close to the time of the peak daily maximum O_3 concentration. Stations were not used with missing hourly averages within two hours of the estimated time of maximum O_3 concentration.

Table 1 summarizes the results of the 154 days of high O_3 concentrations for each station. The variability in the daily maximum O_3 concentration from station to station is large. Ozone monitoring stations that were located in and near intense emission areas (stations 101 through 107, 111, 112, 113, 119, and 120) were less frequently within 80-100 percent of the highest 1-h O_3 in the RAMS than stations outside of these concentrated emission areas. On the other hand, several stations located outside intense emission areas on a number of days had very low maximum concentrations when compared to maximum 1-h O_3 concentrations at other stations. This is indicated in Table 1 by

the large percentage of time stations 117, 123, 124, and 125 were only 20-40 percent of the maximum 1-h O_3 concentration. The fact that stations outside and north of intense emission areas (station 114, 115, 121, and 122) rarely (1 percent of the time) measured concentrations only 20-40 percent of the maximum 1-h O_3 concentration can be attributed to the fact that southerly winds in St. Louis are more frequently associated with high O_3 concentrations than any other wind direction[9]. This implies that stations north of St. Louis were more frequently in the 'urban plume' during days of high O_3 concentrations than those to the south.

The 154-day data set was used in another way. The combination of stations was determined, for a given number of stations, that had the greatest number of days with at least one station's concentration within 80-100 percent of the highest 1-h O_3 concentration in the network. An increase in the given number of stations did not necessarily increase the number of days with at least one station's concentration within 80-100 percent of the maximum 1-h O_3 concentration in the network. For instance, the combination of the three stations 101, 102, and 103 had fewer days within 80-100 percent of the highest O_3 concentration than did the single station 114. In this analysis, the number of stations was increased until at least 95 percent of the high O_3 days could be represented by a subset of the 25 RAMS sites. The process was terminated at 95 percent due to the extensive amount of computer time required to search through all possible combinations of stations.

The results of the search are contained in Table 2. This table illustrates the difficulty in measuring the highest 1-h O_3 concentration in a metropolitan area on days when O_3 concentrations are high. Even the most successful combination of stations required six stations to come within 80-100 percent of the maximum 1-h O_3 concentration on 95 percent of the days when the concentration in the RAMS exceeded 100 ppb. Notice that these six stations were all located outside concentrated emission areas. This does not imply that O_3 concentrations fail to reach high values inside the city. Figure 3 reveals that the absolute maximum 1-h O_3 concentration in urban areas is nearly as high as in rural areas. However the frequency of attaining these high concentrations is considerably less in intense emission areas. This is apparent by the lower average 1-h O_3 concentrations at the urban sites.

Figure 4 contains the 0700-0800 CDT hourly average concentrations of CO, THC, NO, AND NO_2 for all stations in the RAMS. For all pollutants other than O_3, the highest average concentration occurred during 0700-0800 CDT. Unlike the Figure 3 distribution of average O_3 concentrations, Figure 4 shows that for CO, THC, NO and NO_2, no station outside the intense emission areas had higher 1-h average concentrations than stations inside these areas.

MAXIMUM O_3 CONCENTRATIONS AND THE AIR FLOW

At each station the O_3MAXP value was calculated, and related to daily network values of wind direction, speed, and steadiness. The daily network wind speed was the arithmetic average of 24 individual hourly speeds, each of which was the median speed for the 25 stations for each hour. The daily network wind direction was the resultant of 24 individual hourly directions, each of which was the resultant direction for the 25 stations for each hour. The steadiness of the wind was computed using the following expression:

$$S = \frac{\overline{V}_R}{\overline{V}}$$

where $\overline{V}$ is the 24-h arithmetic average wind speed (i.e., the network speed as described above), and $\overline{V}_R$ is the 24-h resultant wind speed (i.e., based on the network direction described above). For each station (other than 101) each resultant wind direction was categorized into an upwind or downwind quadrant that was defined by the relative location of station 101 (the area of high emissions). For example, in Figure 5 if the resultant wind direction was between 310° and 40°, station 105 was considered downwind of station 101, and if the wind was between 130° and 220°, it was classified as an upwind day. The network wind speed was categorized as high (>3.4 m/s), medium (2.5-3.4 m/s), or low (<2.5 m/s). These categories were determined so that each had an equal number of days. At each station the O_3MAXP value was calculated for high and low wind speeds and for days of upwind or downwind flow.

Figures 5 and 6 depict the relationships between the wind (direction and speed, respectively) and the location of the O_3 maximum in the RAMS. As the distance from the intense emission area increases, there is a large difference between upwind and downwind O_3 concentrations, e.g., at the outer sites compared to the inner sites (stations 114 through 125 versus stations 102 through 113). When wind speeds were high (Figure 6) a closer approximation to the network maximum O_3 concentration was generally observed at the stations away from concentrated emission areas than at stations in and close to these source areas. For example, on days with high wind speed the percentages of the network highest O_3 concentrations typically are in the 60's at stations in the vicinity of the high emissions area, but are often in the 70's (some 80's) at more distant stations. Therefore, on days with high O_3 concentrations (>100 ppb) and relatively strong wind speeds the highest O_3 concentrations are not likely to be found near intense emission areas. For low wind speeds there is a better chance that the highest O_3 concentration in a metropolitan area is located closer to the intense emission areas than for high wind speed days. A specific example of this occurred on October 1 and 2, 1976. The lowest wind speeds during the entire period of study were reported on these days and the highest O_3 concentrations were all in or near concentrated emission areas.

The steadiness of the wind was also classified into three categories and related to the O3MAXP values. The steadiness of the wind did not appear to have much significance as an indicator of which type of site would be representative of the maximum O_3 concentration in the RAMS. This was determined by inspection of various contingency tables and t-tests.

HOURLY POLLUTANT CONCENTRATIONS AND THE URBAN PLUME

The network median concentration for each of the five pollutants THC, CO, NO, NO_2, and O_3 was calculated for the four time periods, 0700-1000-1100, 1300-1400, and 1600-1700 CDT, based on all available measurements. These network median concentrations were then subtracted from the corresponding 1-h average concentrations at each station and categorized according to a network wind direction which was determined in the following manner. On each day for each of the four 1-h periods of interest the

network wind direction was classified into one of three sectors, 30°-150° (E), 150°-270° (SSW), and 270-30° (NNW). For each hourly period of interest, the network wind direction was based on up to 125 hourly wind measurements (when none were missing), the hour of interest and the four hours previous at each of the 25 RAMS sites. To be classified into one of the three specified direction sectors 75 percent of the wind directions had to fall within a single sector with a speed exceeding 1.5 m/s. For each pollutant and time period of interest, the concentration deviations from the network median were then averaged over all the days of identical wind direction sectors.

The decision to use five hours of data and a minimum wind speed of 1.5 m/s was based on the travel time necessary for the pollution emitted from the concentrated emission areas to reach the more distant RAMS sites. The subtraction of the network median concentration for each of the three sectors normalizes the concentrations of one sector with respect to the concentrations of the two other sectors. This alleviates the first major difficulty in understanding some of the more subtle effects of the city on pollutant patterns. For example, due to the weather regimes that affect St. Louis, O_3 concentrations over the entire network are generally higher with southerly winds than with northerly winds[9]. Because of this it is not uncommon to observe higher O_3 concentrations at sites south of the city on days of upwind flow as opposed to downwind flow.

A second subtraction (or normalization) was required in order to normalize concentrations within a sector. An example of why this was necessary can easily be given. Normally CO concentrations are much higher in and near the city as opposed to stations farther removed. When the network median CO concentration is subtracted from each station's CO concentration, the urban stations consistently have concentrations above the median, regardless of wind direction, and the rural sites consistently have less than the median CO concentration. In order to overcome this effect the deviation of a station's pollutant concentration from the network median concentration, after being averaged for each one of the three wind direction sectors, was subtracted by the average of

the deviations from the network median concentration of the two other wind sectors. The net result is a normalized field of concentrations representing the mean departure from the normalized concentration fields of the two other wind sectors. These concentrations are plotted on Figure 7 for the four time periods of interest.

Figure 7a depicts higher THC concentrations downwind of the city for all time periods with winds from the SSE-W sector*. Figure 7b illustrates that the urban plume concept, at least for mean conditions, applies even for a pollutant such as CO. This pollutant, which has considerable spatial variability, shows consistent patterns of higher concentrations downwind of source areas. Figure 7c depicts higher concentrations of NO downwind of the city during the early morning and late afternoon, but during other times around mid-day such a pattern is not readily apparent. This is at least partially attributable to the very low NO concentrations, not infrequently near the noise level of the instrument, that occur during these hours. Figure 7d is very much similar to Figures 7a and 7b in that higher than normal NO_2 concentrations occur downwind of the city's intense emissions areas throughout the day. In Figure 7e, O_3 concentrations were lower than normal downwind of the city until approximately 1000 CDT, but after 1000 CDT, O_3 concentrations were higher than normal in areas downwind of the city. The lower O_3 concentrations that occur before 1000 CDT are attributed to the scavenging of O_3 by NO since patterns of NO and O_3 were similar but opposite in sign for the period 0700-0800 CDT. Apparently, after 1000 CDT the energy of the solar radiation and the subsequent reaction time necessary to provide O_3 are sufficient to overcome this phenomenon. The changeover time from lower than normal O_3 concentrations downwind of the city to higher than normal concentrations during 1000-1100 CDT is confirmed by Figure 7f for other wind directions.

* Figures 7a and 7e do not contain the pollutant patterns for the W-NNE or the NNE-SSE wind sectors since these wind directions had nearly the same type of patterns as did the SSE-W wind sector.

CLUSTER ANALYSIS OF POLLUTANT CONCENTRATIONS

In order to understand the spatial variability of pollutant concentrations, it is essential to investigate correlation between concentrations at various stations. These correlations can then be reconciled in a systematic approach through the use of cluster analysis.

Correlation coefficients were calculated for each pollutant at each pair of stations. The correlations were calculated for one of the four hours 0700-0800, 1000-1100, 1300-1400, or 1600-1700 CDT, depending upon the hour which had the highest network average concentration. In order to avoid inflated correlation coefficients due to the yearly oscillation of pollutant concentrations, the actual concentrations at each site were subtracted by their seasonal means before the correlations were run. Three seasons were formed, consisting of the months April and May, June through August, and September and October. Since the subtractions effectively removed the yearly oscillation these deviations from seasonal means were then used to compute the correlation coefficients.

Cluster analysis was applied to the matrix of correlation coefficients produced by the interstation correlations. Cluster analysis is achieved by combining stations into groups or clusters. Stations are combined into a cluster based on the correlation coefficients between the various stations. Initially, each station can be thought of as a cluster unto itself (a correlation coefficient of unity). At each step in the clustering technique all possible combinations of existing clusters are considered. Only the two clusters which have the highest average correlation coefficients, based on the station-to-station correlation coefficients, are combined into a new cluster. Once a cluster is formed it cannot be broken apart. In this manner clustering continues until all the stations are combined into one cluster. Since there are 25 stations this process requires 24 steps to complete.

Cluster analysis does not require any assumption regarding the statistical significance of the correlation coefficients. The use of average correlation coefficients can be justified because the clustering technique initially combines stations into a cluster which have "high" correlation

coefficients and works down to "low" correlation coefficients by the final step. In this manner averages of correlation coefficients are considered for only the pairs of stations which have similar correlation coefficients (i.e. high, medium, or low). The problem of averaging or combining correlation coefficients which differ significantly[10] is then sidestepped.

Figure 8 contains the results of cluster analysis for the various pollutants. For O_3 the cluster analysis is also included for the maximum daily 1-h O_3 concentration when at least one station reports a concentration in excess of 100 ppb. Each of the 24 steps for each pollutant is contained in Figure 8. A new cluster is formed at each step represented by the stations included within the boundaries of the solid block containing correlation coefficients. The first correlation coefficient is the average correlation coefficient of all pairs of stations between the previously distinct clusters. The second correlation coefficient listed is the minimum correlation coefficient of any pair of stations between the previously distinct clusters.

Figure 8a is for the hour with the highest network average concentration of CO. The figure reveals a poor correlation between the RAMS sites with respect to CO concentrations. By the sixth step the similarities of any two clusters is so low that the average variance between any two clusters is less than 50 percent*. An inspection of the stations included in the clusters indicates that they were formed on the basis of local emissions around each station, the distance separating the stations, and their orientation with respect to the central urban area of concentrated CO emissions. Some clusters were created when there was a large disparity in the local emissions surrounding specific stations in the clusters. This only occurred when stations were separated by relatively short distances (see step 6, Figure 8a). In other clusters (steps 2 and 16) the local CO emission around the station and the orientation of the station with respect to major emission

* This calculation and subsequent calculations in this section regarding the average variance explained by one cluster for another cluster were calculated using the average of the square of the correlation coefficients between stations. The square of the average correlation coefficients shown in Figure 8 turns out to be a good approximation to the variance.

areas was more important than the distance separating the station. The formation of a cluster which is more difficult to explain occurs at step 18. It is unclear why station 105, in an area of intense CO emission, does not cluster with other nearby stations in concentrated emission areas before it clusters with stations 104 and 120. One explanation is that station 120 is affected more by a local interstate highway approximately 2 km to the west (Figure 1) than by the large central area of high emission further east. This implies that stations 104, 105, and 120 were oriented east of important emission areas.

The emission pattern for THC is remarkably similar to the pattern for CO, and cluster analyses for these two pollutants also have similarities. For example, the cluster formed at step 11 in Figure 8b is almost identical to the cluster formed at step 9 of Figure 8a Clusters formed at steps 7, 8, 12, and 16 of Figure 8b have identical counterparts on Figure 8a The correlation between stations is slightly higher for THC than CO, but it is also low by the final step. In fact, if a subset of stations from the RAMS were desired to explain an average of 50 percent of the variance in THC concentrations, no less than 16 stations would be required (see step 9, Figure 8b).

Cluster analysis for NO and NO_2 (Figures 8d and 8e) also reveals poor correlation coefficients, particularly for NO. A subset of the RAMS would require 24 stations for NO and 16 stations for NO_2 in order to explain an average of at least 50 percent of the day-to-day variance of the concentrations at each station from the original network.

The correlations for O_3 concentrations during 1300-1400 CDT are quite high as indicated in Figure 8e. The clusters formed in this analysis are closely tied to the orientation of the stations with repsect to major emissions areas as well as the distance between stations. Clusters formed by this secondary pollutant are highly related. Unlike the clusters formed by the four other pollutants, these clusters suggest that stations are representative of areas rather than points. In Figure 8e step 10 suggests that a single station could be used to replace most of the stations in and near the major source areas of urban St. Louis. Even in areas outside the city there are a number of clusters with sufficiently high correlation

coefficients such that a single station could be used in place of two, three, or even four stations. However, in order to employ one station as a surrogate for another station the average difference of the O_3 concentrations between the stations during 1300-1400 CDT must be added to the surrogate station's concentration during the same hour.

Figure 8f is important in light of the NAAQS emphasis on maximum 1-h O_3 concentrations. Maximum 1-h O_3 concentrations have higher correlation coefficients between stations than average hourly correlations for the other pollutants, but these correlations fall short in comparison to average hourly concentrations of O_3 during 1300-1400 CDT, the hour with the network average highest concentration (see Figure 8e). The formation of clusters in Figure 8f were closely related to the distance between stations. This implies that when maximum O_3 concentrations were relatively high (low) at specific stations an area of high (low) maximum O_3 concentrations surrounded this site, but the correlation coefficients indicate that this area was not large in size. A similar result is found in the analysis of a few days of data during an O_3 episode in St. Louis[11].

The information contained in Figure 8f can be used with Table 1 to expand the list of stations in Table 2 that have concentrations 80-100 percent of the daily highest 1-h O_3 concentration in the RAMS. The six stations in Table 2; 114, 115, 118, 122, 124, and 125 are in distinct clusters at step 18 in Figure 8f (initially each station is a cluster unto itself). At step 17 station 109 is not included in any of the clusters with these six stations. Table 1 indicates that station 109 has the highest percentage of days within 80-100 percent of the maximum 1-h O_3 concentration of the 19 stations <u>not</u> listed in Table 2. The seven stations 109, 114, 115, 118, 122, 124, and 125 (i.e., station 109 plus the six stations in Table 2) were within 80-100 percent of the maximum O_3 concentration in the RAMS 97 percent of the time. Figure 8f indicates that at step 16, station 120 is not included in any of the clusters of these seven stations. Furthermore, in Table 1, station 120 has the highest percentage of days within 80-100 percent of the highest 1-h O_3 concentrations of the remaining stations. With this addition, at least one of the eight stations, 109, 114, 115, 118, 120, 124, and 125, was within 80-100 percent of the maximum 1-h O_3 concentration in the RAMS on <u>each</u> of the 154 days.

SECTION 5

SUMMARY

Stations that monitored O_3 concentrations in and near intense emission areas were less frequently within 80-100 percent of the highest 1-h O_3 concentration in the RAMS on days when the concentration exceeded 100 ppb than stations away from these areas. On the other hand, several stations located south and upwind of the city had very low 1-h maximum O_3 concentrations compared with other stations. A total of eight stations, all located outside of intense emission areas, could have been used in place of 25 to detect (within 80-100 percentage) the 1-h network maximum O_3 concentration when it exceeded 100 ppb. Mean maximum concentrations of all the other pollutants studied were located inside areas of intense emissions.

On days when the 1-h O_3 concentration exceeded 100 ppb and relatively strong wind speeds (>3.4 m/s) were observed, the highest O_3 concentration in the network was rarely located near intense emission areas. For low wind speeds (<2.5 m/s) the highest O_3 concentrations were more often located closer to the intense emission areas.

The scavenging of O_3 by NO resulted in lower than normal 1-h average O_3 concentrations downwind of the city's intense emission areas before 1000 CDT. After 1000 CDT O_3 concentrations downwind of the city averaged higher than normal. The concept of an urban plume was verified for mean concentrations for all the primary pollutants, even for the highly variable pollutant CO.

Station-to-station correlations of the primary pollutants was generally lower than 0.70, except for some closely spaced stations, particularly those in intense emission areas. Station-to-station correlations of O_3 were significantly higher than those of the primary pollutants. In fact, a single station frequently could be used in place of several nearby stations with little loss of information. This was particularly true for the stations located in and around the urban area of concentrated emissions.

REFERENCES

1. Federal Register (Thursday, June 22,1978). Photochemical Oxidants; Measurement of Ozone in the Atmosphere; Requirements for Preparation, Adoption, and Submittal of Implementation Plans. Vol. 43, No. 121.

2. Larsen, R. I. A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards. AP-89, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina. 1971. 56 pp.

3. Staff Summary Report. Alternate Forms of the Ambient Air Quality Standard for Photochemical Oxidants. U.S. Environmental Protection Agency. Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. 1978. 20 pp.

4. McClenny, W. A. and L. W. Chaney. Pollutant Variability in the Regional Air Pollution Study. J. Air Pollut. Control Assoc. 28, 693-696, 1978.

5. Myers, R. L. and Reagan, J. A. The Regional Air Monitoring System, St. Louis, Missouri, U. S. A. Conf. on environmental sensing and assessment proceedings: CAT, IEEE No. 75-CH 1004-ICESA, Las Vegas Nevada. 1975. pp 8-6, 1-9.

6. Pooler, F. J. Network Requirements for the St. Louis RAPS. J. Air Pollut. Control Assoc. 24, 228-231. 1974.

7. Jurgens, R. B. and Rhodes, R. C. Quality Assurance and Data Validation for the Regional Air Monitoring System of the St. Louis Regional Air Pollution Study. Conf. on environmental modeling and simulation proceedings: EPA 600/9-76-016, Cincinnati, Ohio, 1976. 730-734.

8. Littman, F. E. Regional Air Pollution Study Emission Inventory Summarization. U. S. Environmental Protection Agency, Research Triangle Park, North Carolina. 1979. 92 pp.

9. Karl, T. R. and G. A. DeMarrais. Meteorological Conditions Conducive to High Levels of Ozone. In: Report of International Conference on Photochemical Oxidant and Its Control, EPA 600/3-77-001a Sept. 1976 Research Triangle Park, North Carolina. 1977. pp. 75-88.

10. Brooks, C. P. and N. Carruthers. Handbook of Statistical Methods in Meteorology. Her Majesty's Stationary Office, London, 1953. pp. 222-224.

11. Ludwig, F. L. and E. Shelar. Site Selection for the Monitoring of Photochemical Air Pollutants. EPA-450/3-78-013, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina. 1978. 103 pp.

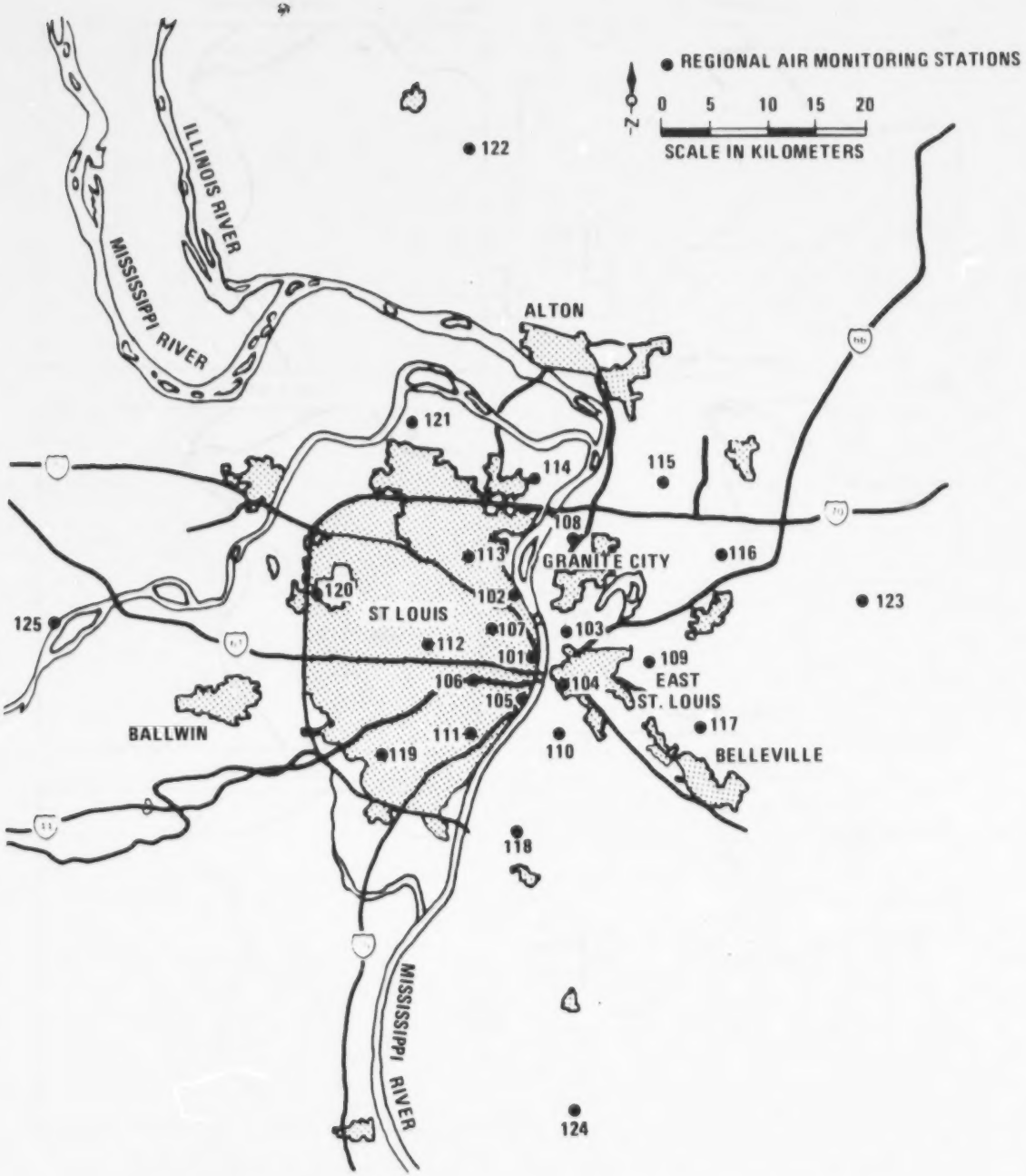

Figure 1. Stations of the Regional Air Monitoring System.

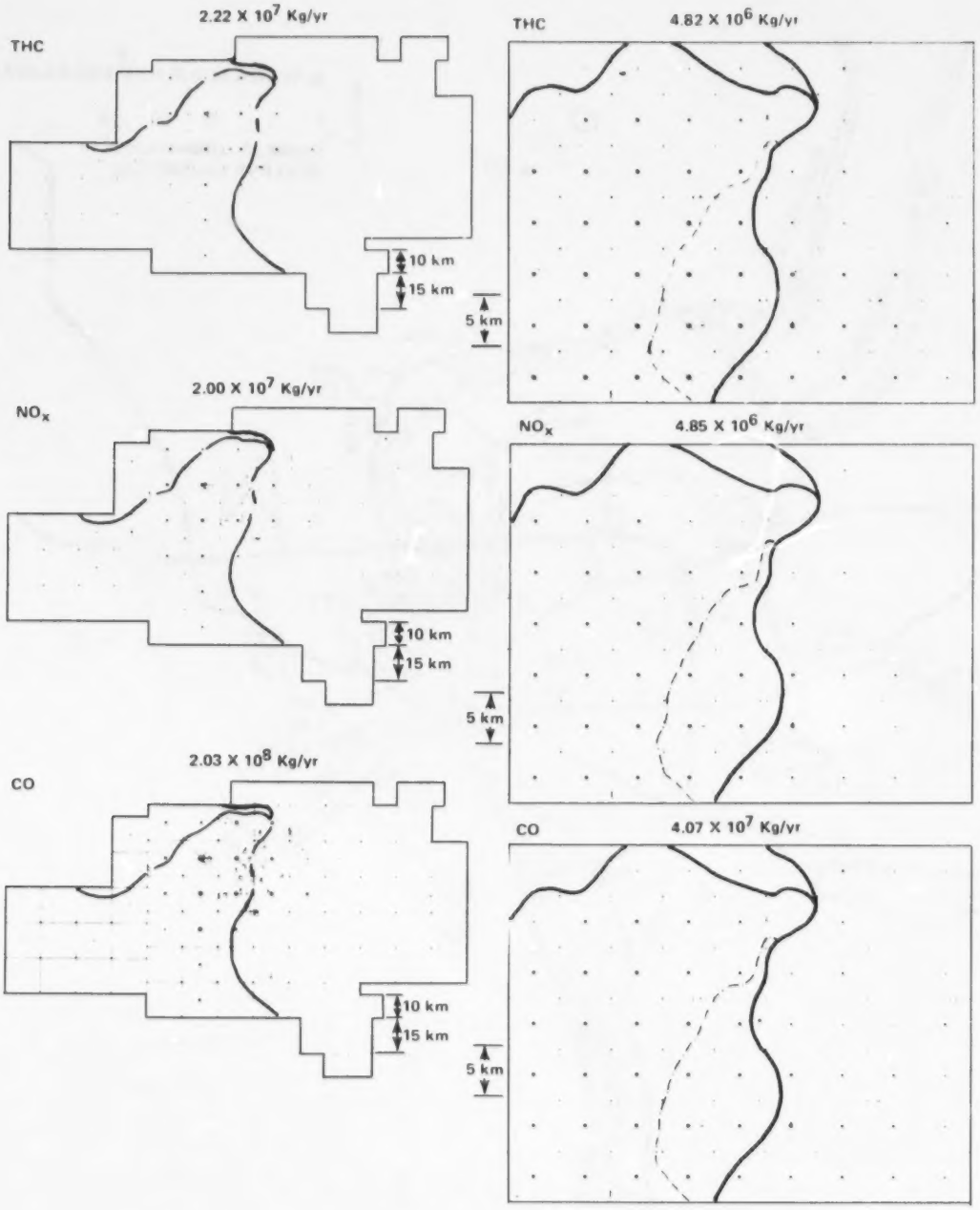

Figure 2. Annual total emissions for the St. Louis area. Diagrams on the right show emissions for 5-km grid squares in the more intense emission areas, while diagrams on the left are for 10- and 15-km girds covering the entire region. Emissions are directly proportional to the area covered by each asterisk; emissions are specified for the largest asterisk in each of the six diagrams.

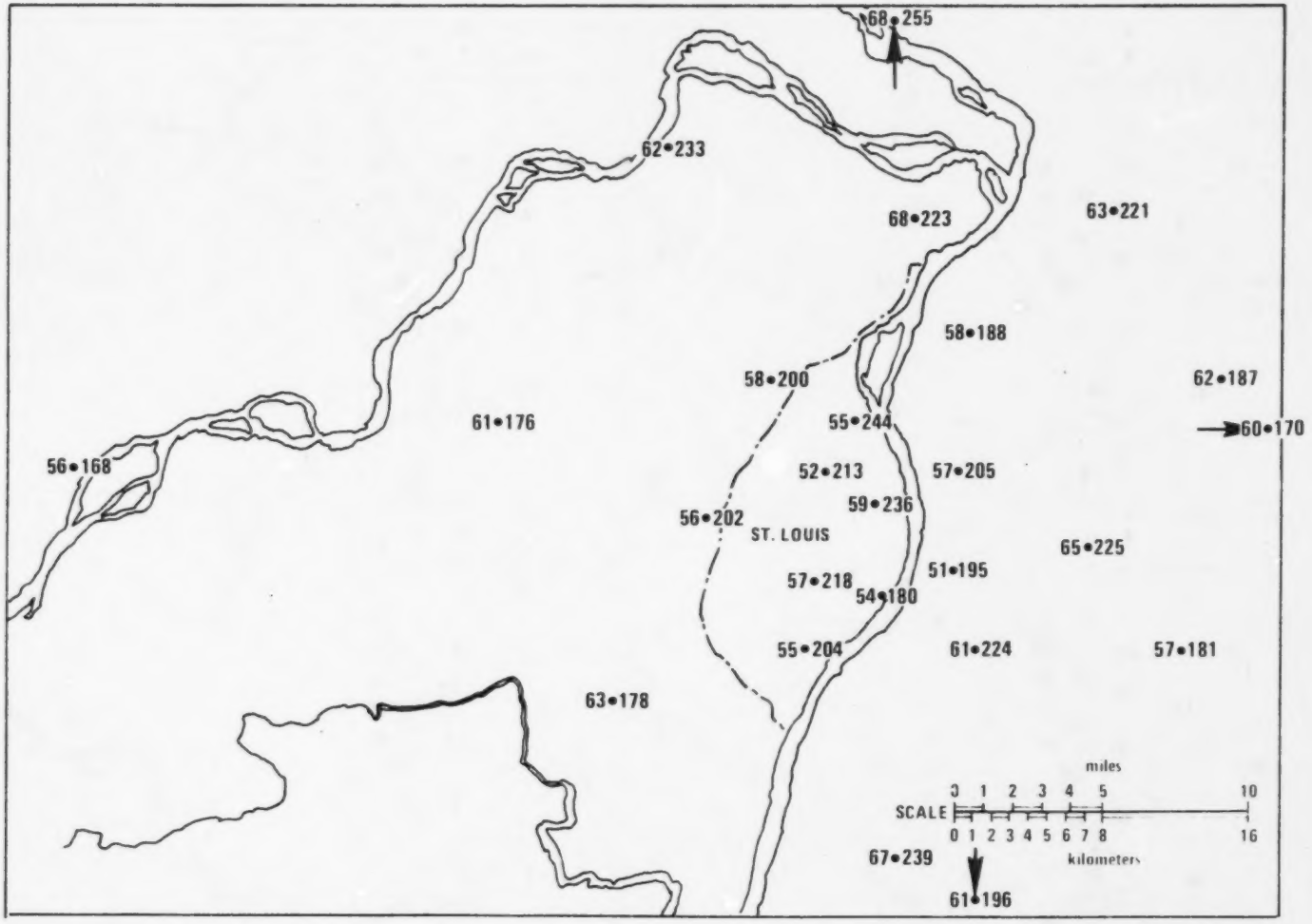

Figure 3. 1-h average 0_3 concentration and the highest 1-h 0_3 concentration (ppb) between 1300 and 1400 CDT for the study period.

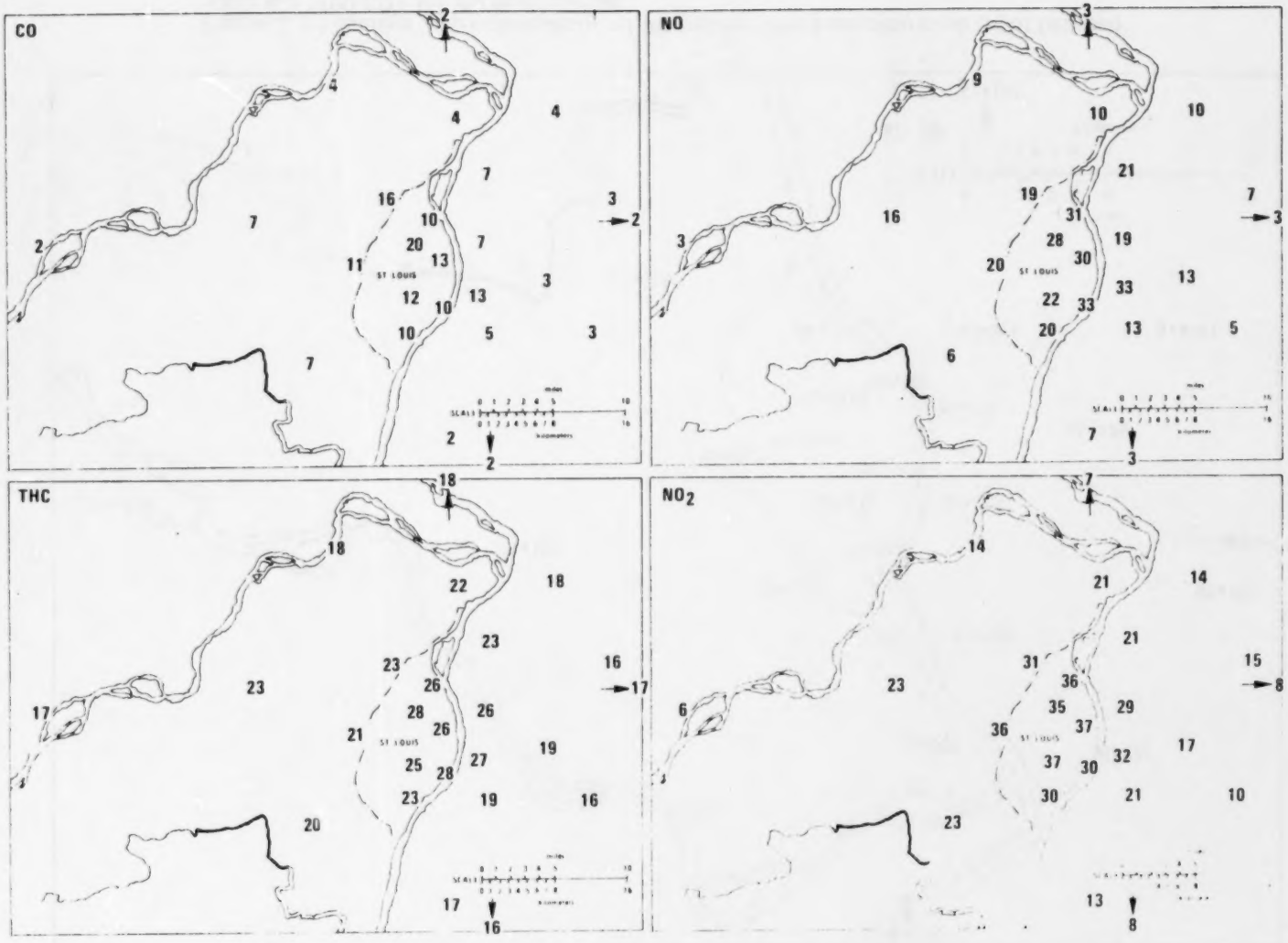

Figure 4. Hourly average concentrations of THC (ppm X 10^{-1}), CO (ppm X 10^{-1}), NO (ppb), and NO_2 (ppb) during 0700-0800 CDT.

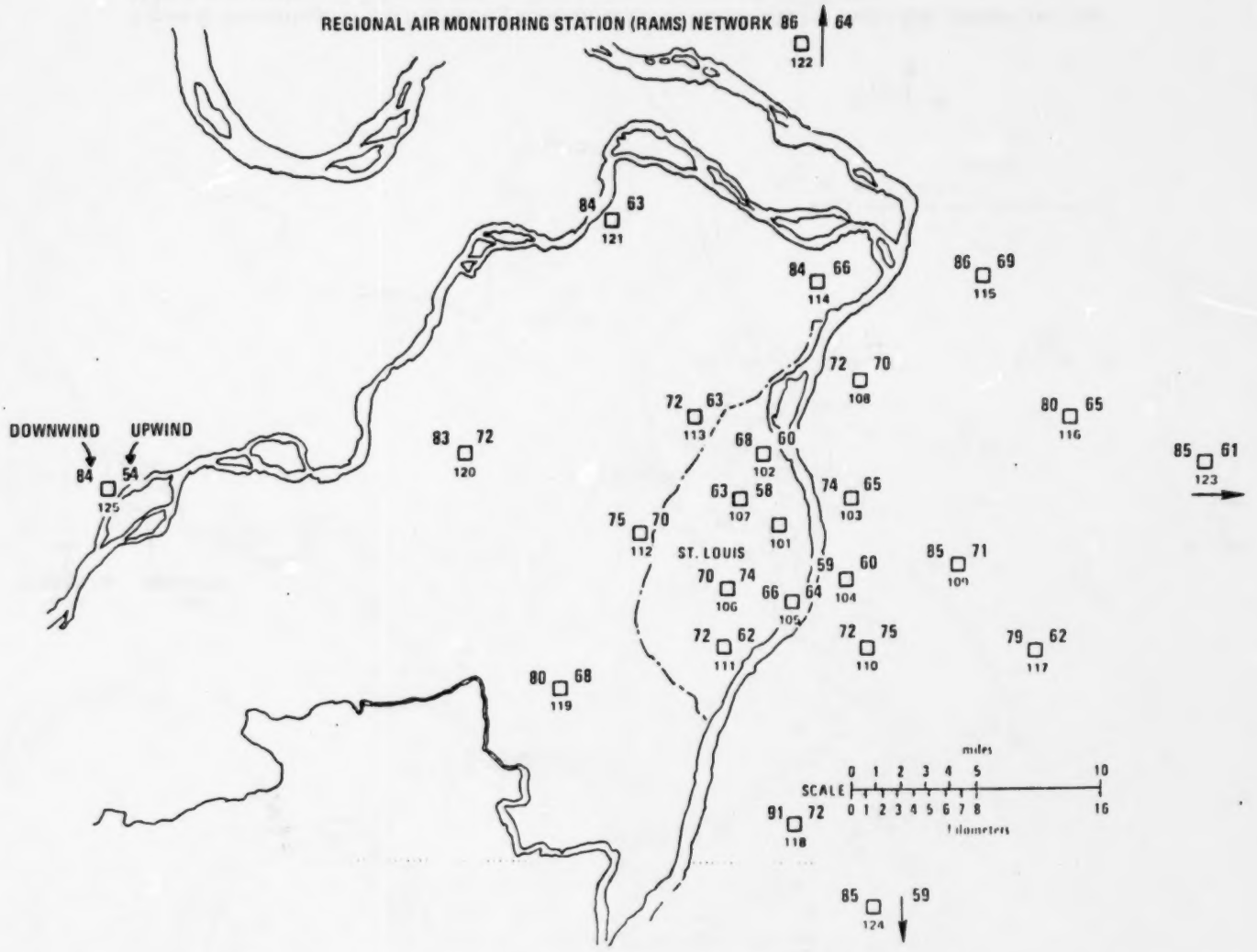

Figure 5. Percentage of the highest O_3 concentration in the RAMS for individual stations on days of downwind and upwind flow.

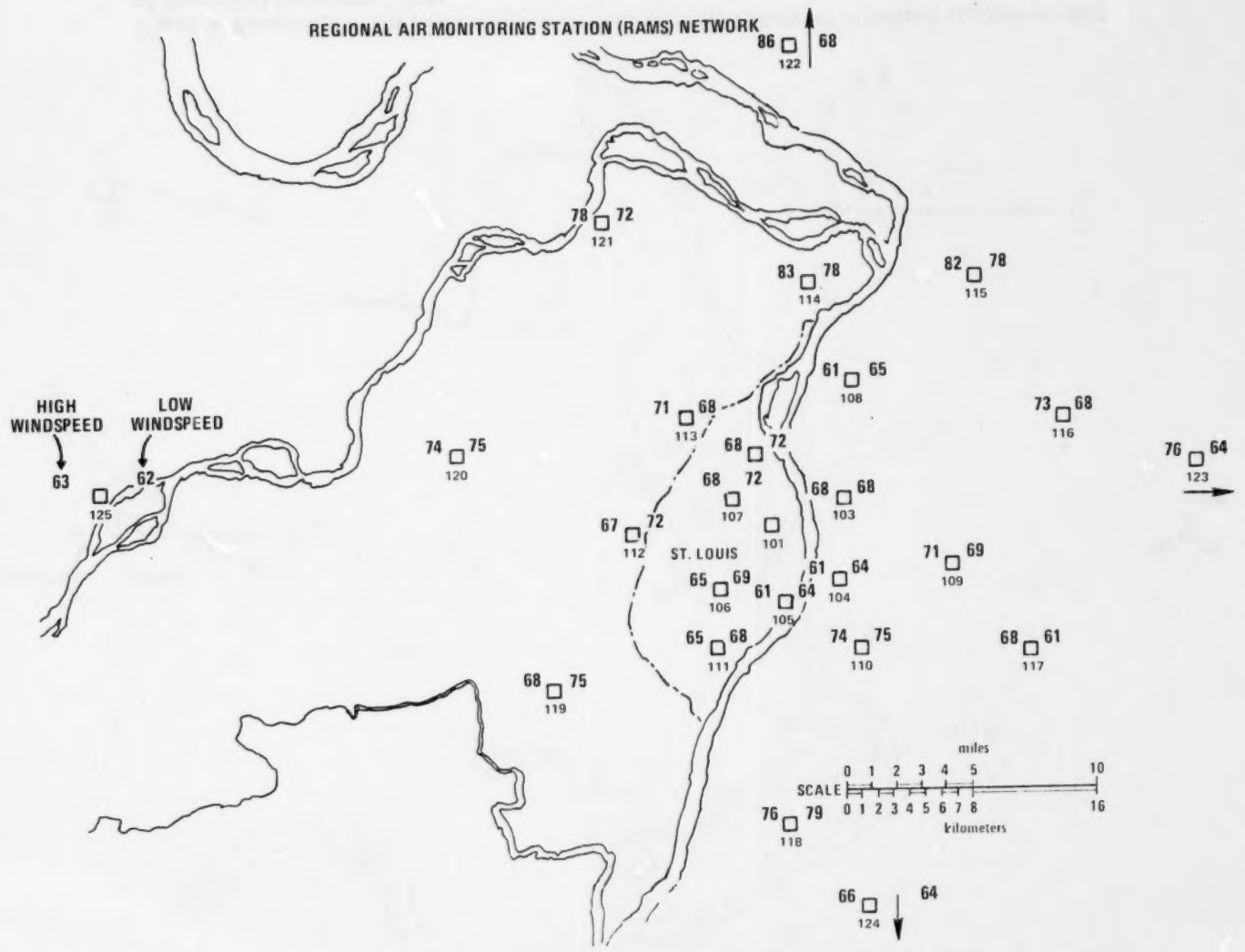

Figure 6. Percentage of the highest O_3 concentration in the RAMS for individual stations on days of high wind speeds (>3.4 m/s) and on days of low wind speeds (<2.5 m/s).

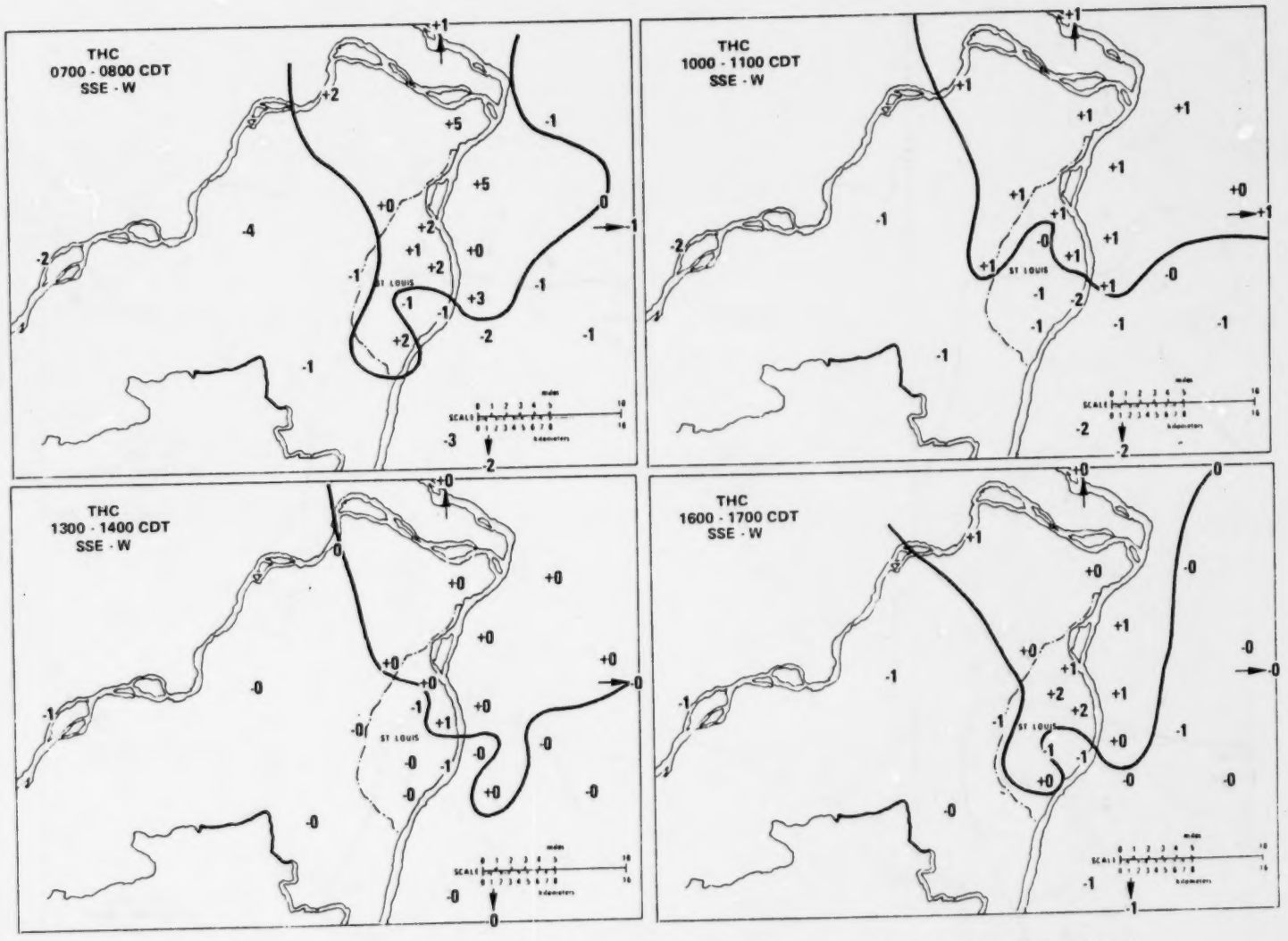

Figure 7a. Normalized concentrations for various pollutants. 0_3, NO, and NO_2 concentrations are in ppb. THC and CO concentrations are in ppm $\times 10^{-1}$. The zero isopleth is indicated on each map.

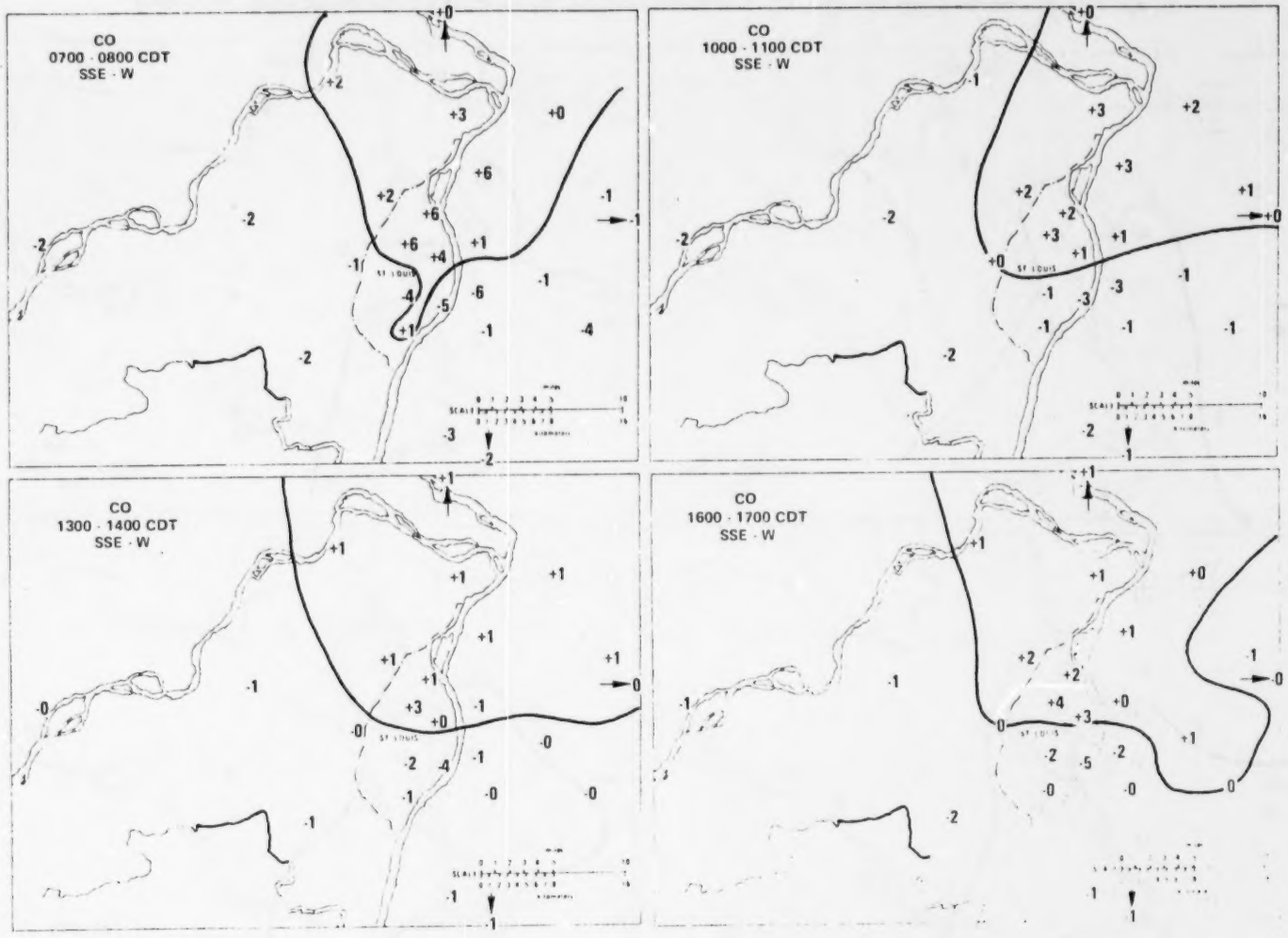

Figure 7b. (Same as figure legend 7a.).

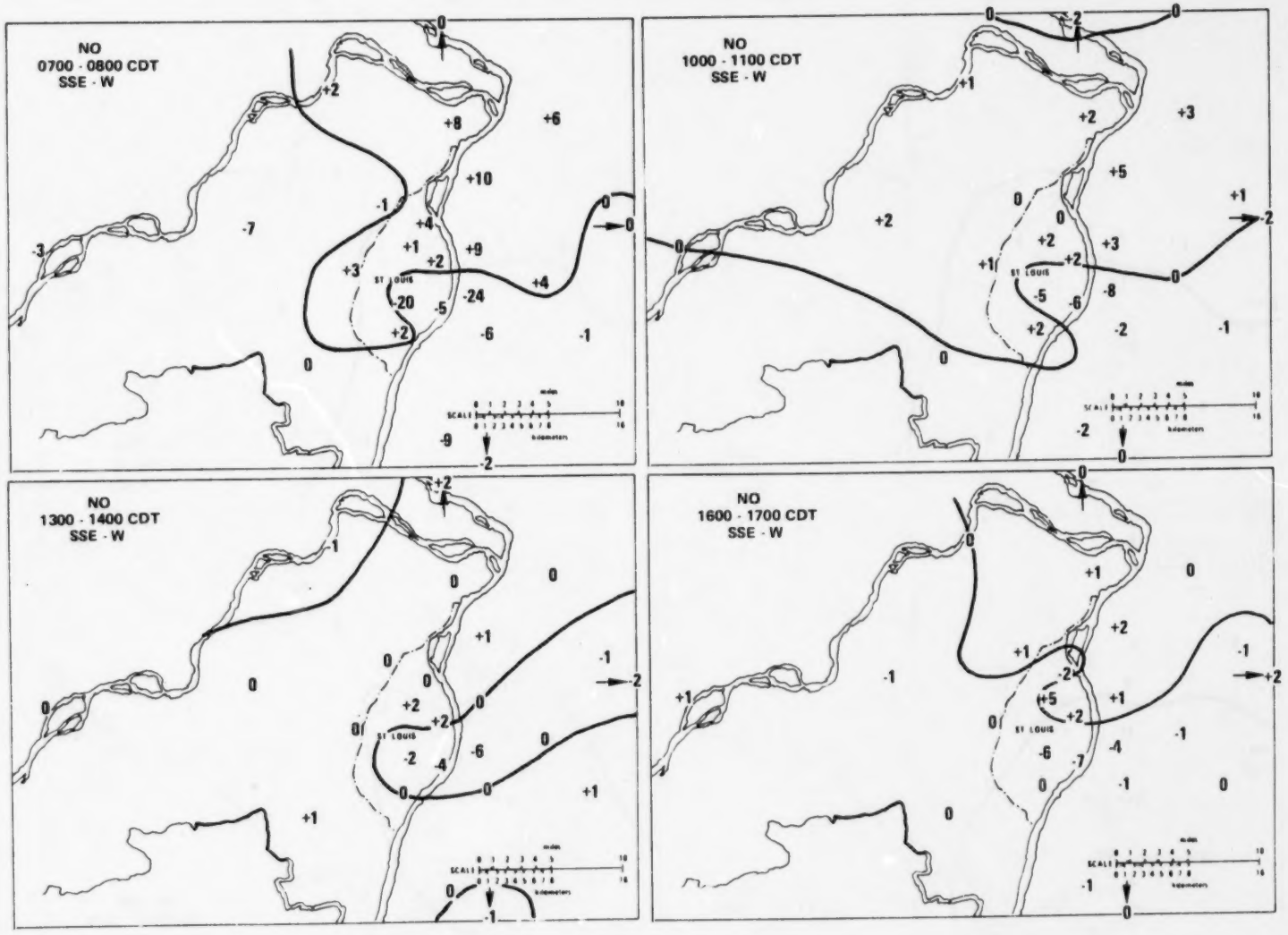

Figure 7c. (Same as figure legend 7a.).

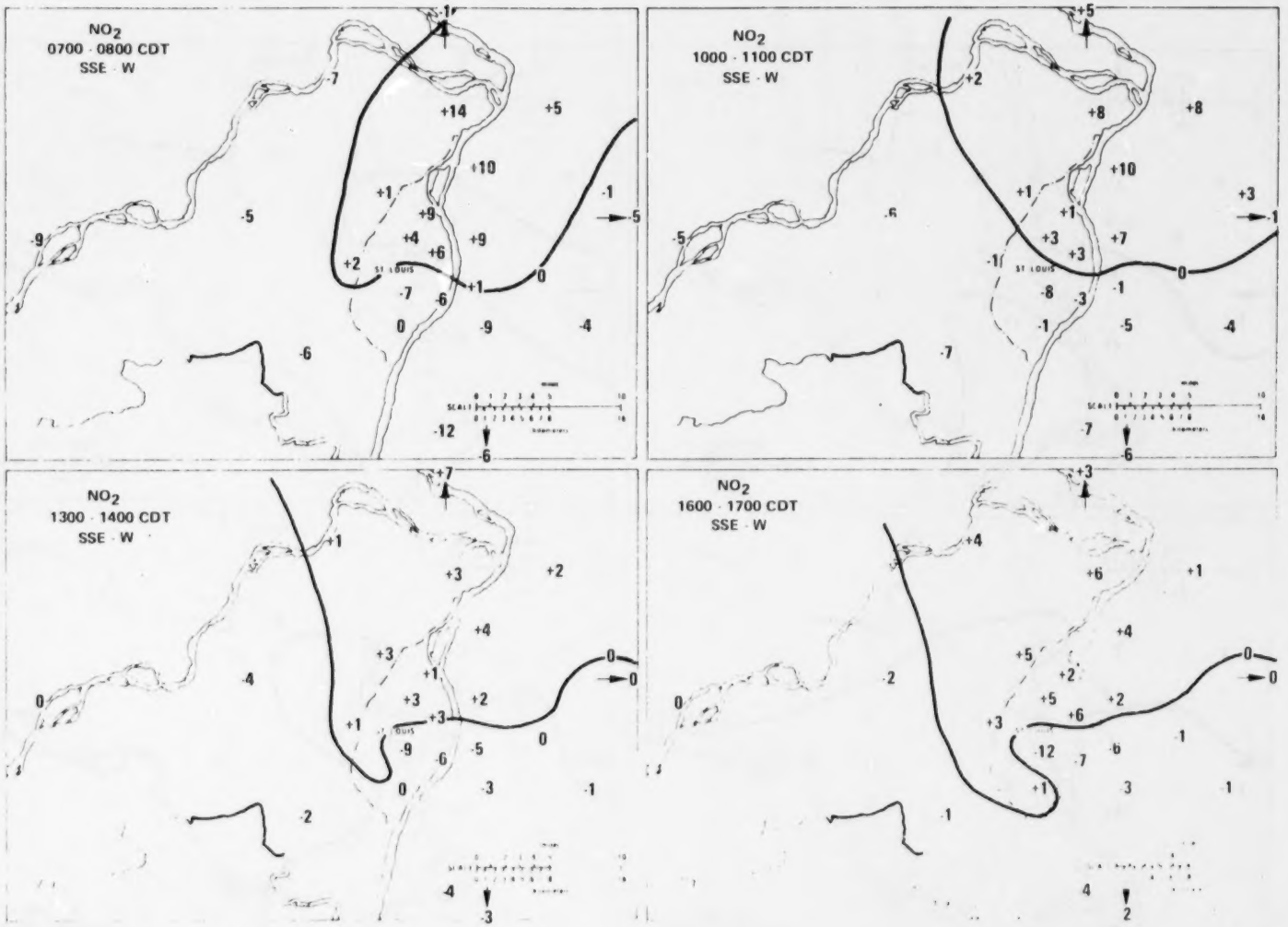

Figure 7d. (Same as figure legend 7a.).

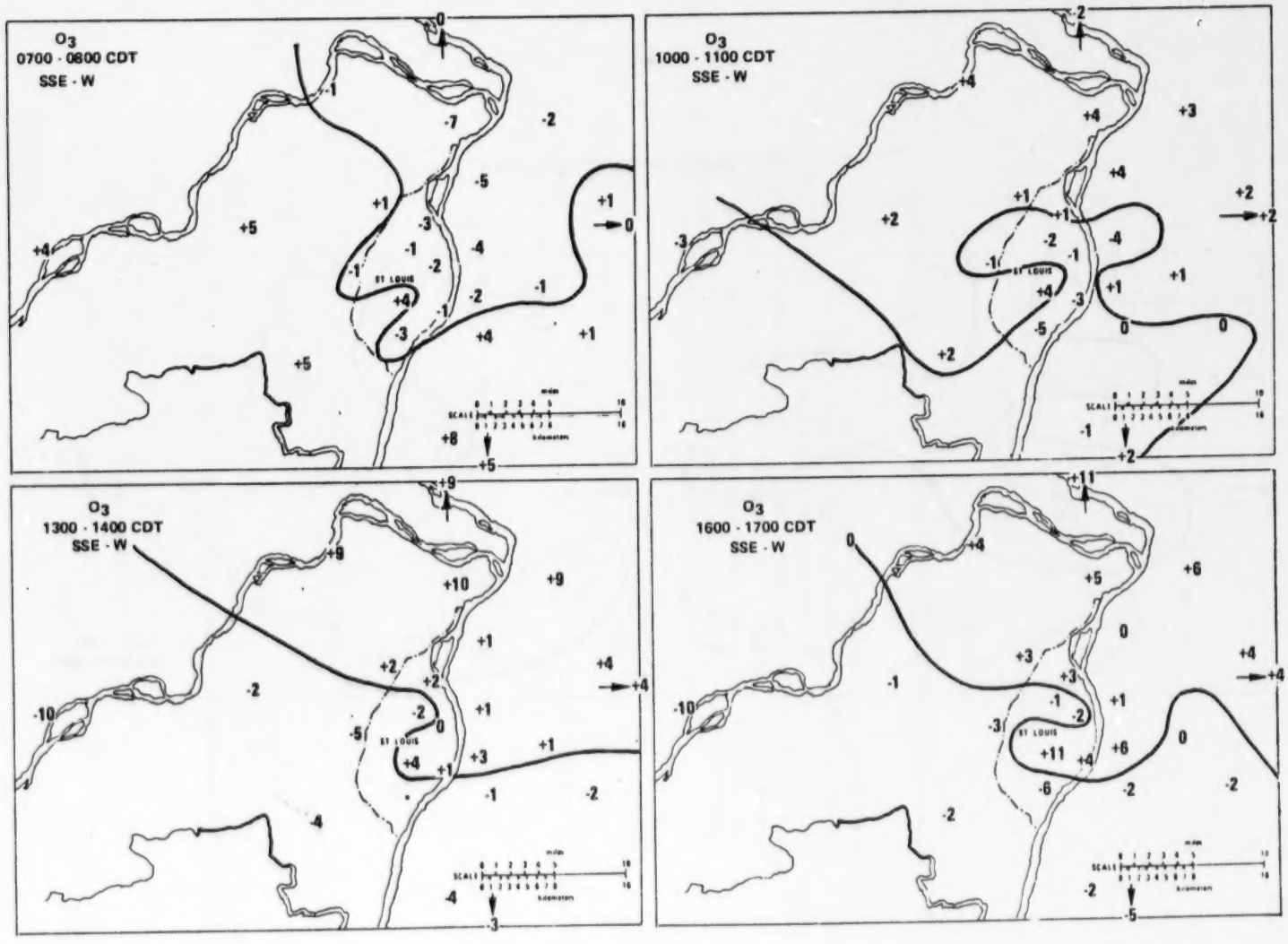

Figure 7e. (Same as figure legend 7a).

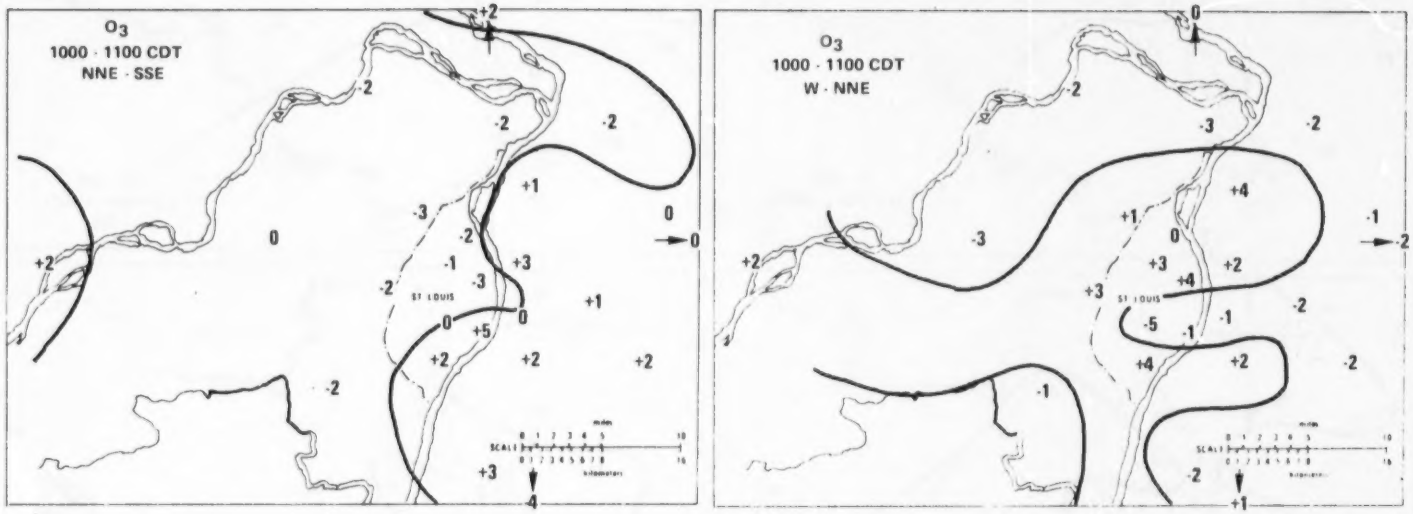

Figure 7f. (Same as legend 7a.).

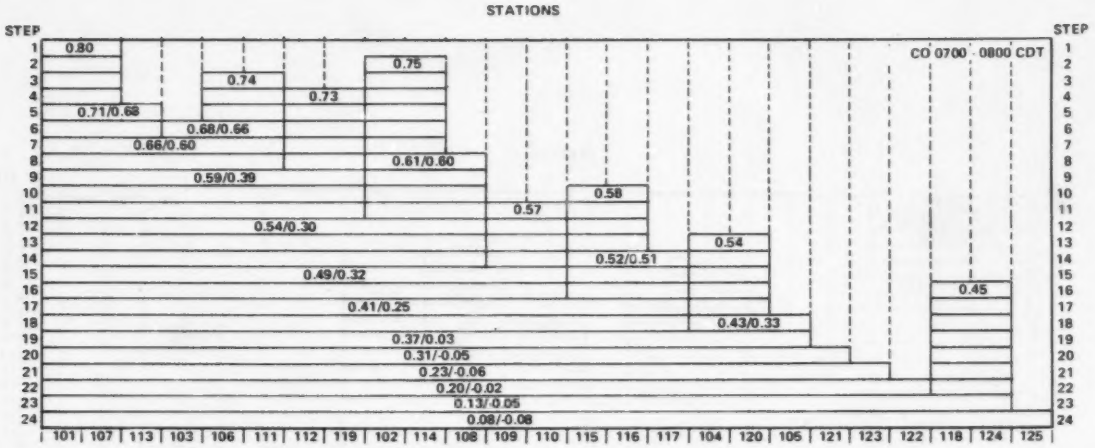

Figure 8a. Diagram for cluster analysis of hourly pollutant concentrations and maximum O_3 concentrations. The first correlation coefficient inside each block represents the similarity of the two clusters which are combined into a single cluster. The second correlation coefficient (or first if the second is absent) is the minimum correlation coefficient between any two stations of the previously distinct clusters. The hour for which each cluster analysis is presented is that for which the network average concentration was highest. See text for further details.

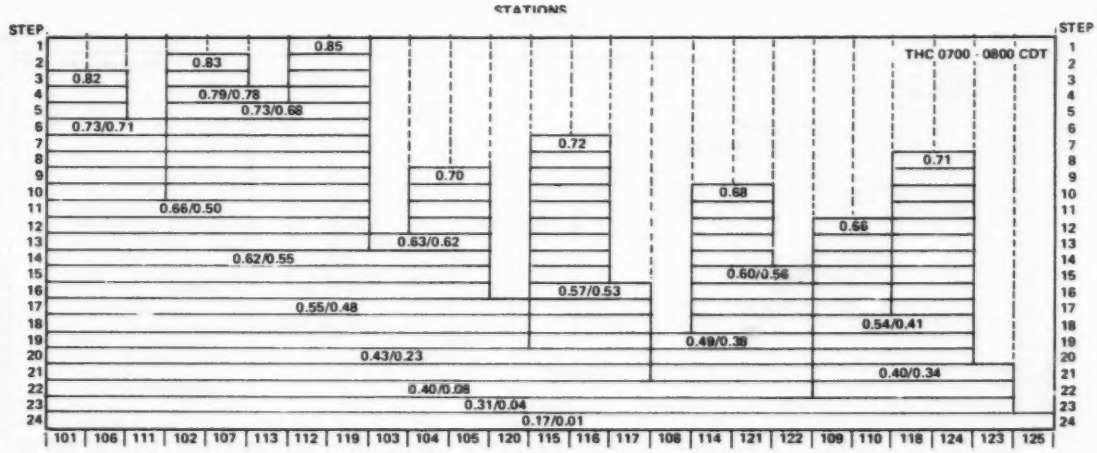

Figure 8b. (Same as figure legend 8a.).

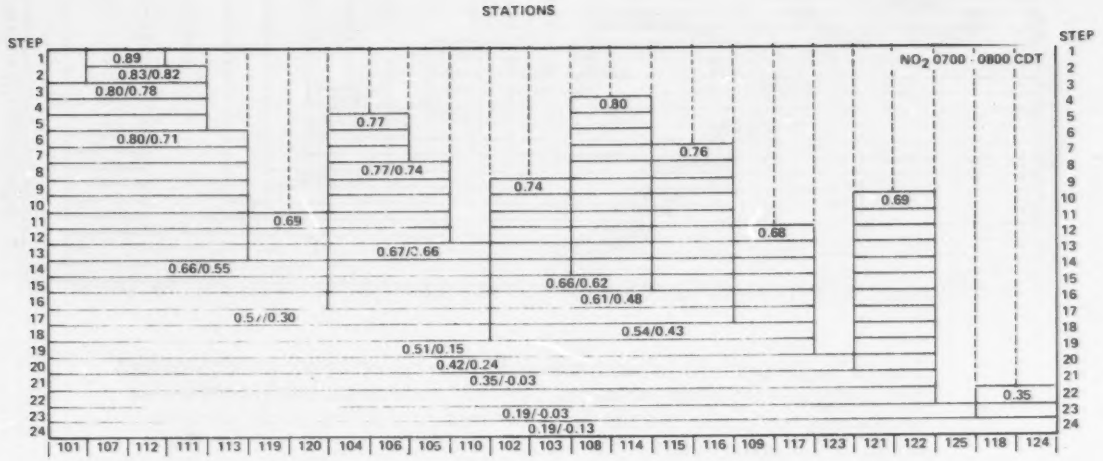

Figure 8c. (Same as figure legend 8a.).

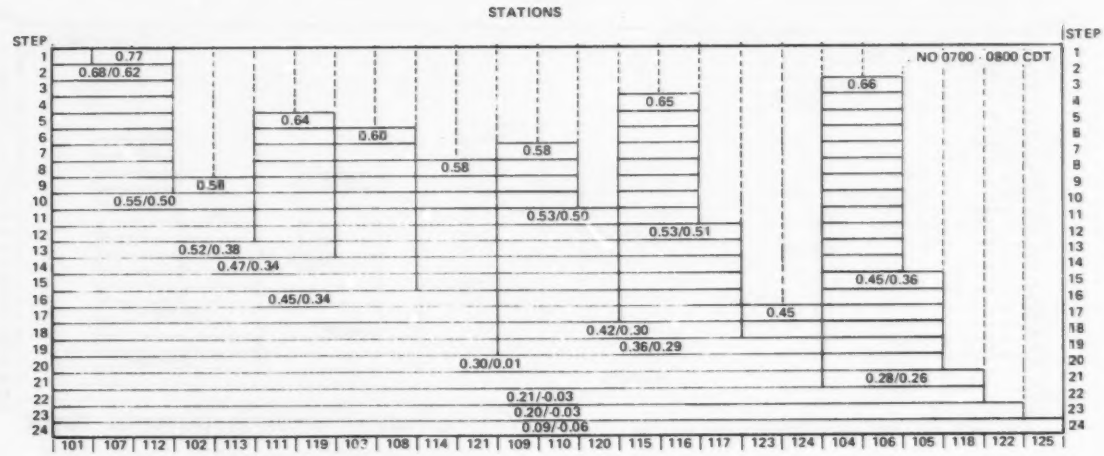

Figure 8d. (Same as figure legend 8a.).

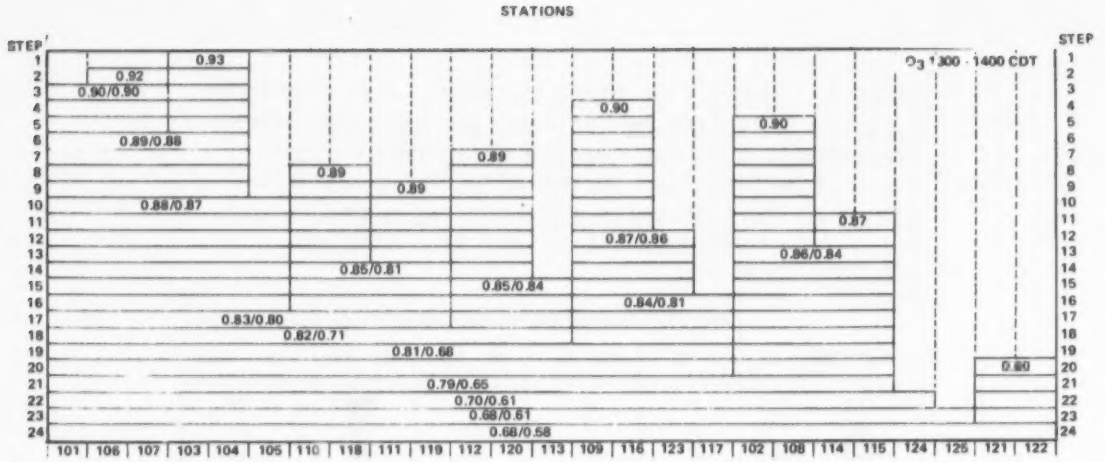

Figure 8e. (Same as figure legend 8a.).

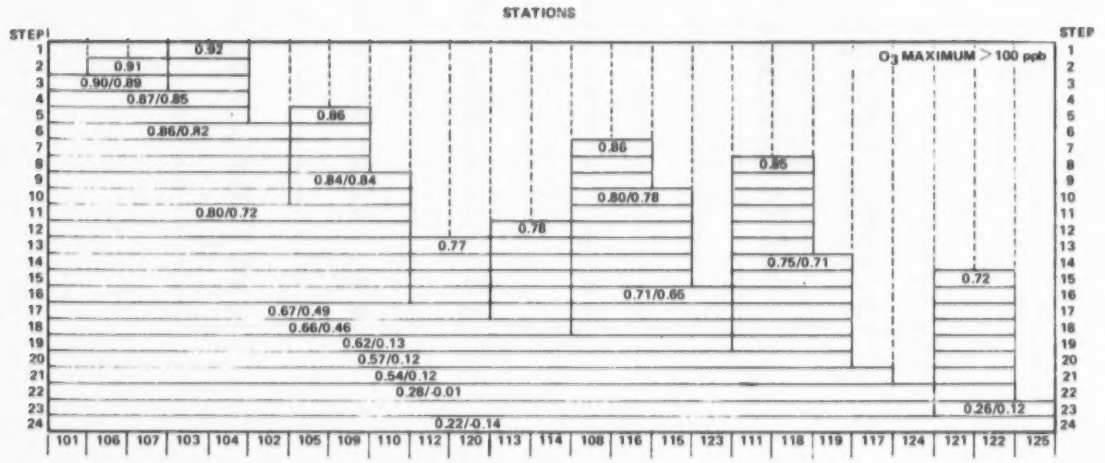

Figure 8f. (Same as figure legend 8a.).

TABLE 1. PERCENT OF DAILY MAXIMUM 1-H O_3 CONCENTRATIONS AT EACH STATION THAT ARE WITHIN GIVEN PERCENTAGE INTERVALS OF THE CORRESPONDING DAILY HIGHEST 1-H O_3 CONCENTRATION IN THE NETWORK. N IS THE NUMBER OF DAYS OF VALID DATA FOR THE STATIONS WHEN THE 1-H O_3 CONCENTRATION AT ANY RAMS STATION EXCEEDED 100 PPB.

STATION	N	PERCENTAGE OF NETWORK HIGHEST CONCENTRATION				
		0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
101	123	0	2	19	54	25
102	132	0	3	25	51	21
103	105	0	4	20	60	16
104	119	0	4	45	44	7
105	123	0	6	22	41	11
106	107	0	5	11	59	25
107	128	0	5	36	52	7
108	130	0	4	24	45	27
109	136	0	3	14	46	37
110	131	0	2	12	56	30
111	110	0	4	27	57	12
112	117	0	1	24	53	22
113	136	0	2	24	53	21
114	128	0	1	11	30	58
115	94	0	1	7	52	40
116	120	0	2	22	52	24
117	122	0	7	30	45	18
118	115	0	2	8	47	43
119	139	0	2	18	44	24
120	88	0	0	11	54	35
121	119	0	1	12	44	43
122	137	0	1	15	65	49
123	66	0	5	35	31	29
124	113	0	9	32	34	25
125	95	0	8	33	36	23

Table 2. Combinations of stations whose concentrations are most frequently within 80-100% of the highest daily maximum 1-H O_3 concentration in the network (for days with maximum concentration above 100 ppb). N is the total number of days above 100 ppb and n is the number of days which were within 80-100% of these high concentrations at the indicated stations.

NUMBER OF STATIONS	STATIONS	n	N	n/N, percent
1	114	74	154	48
2	114, 122	106	154	69
3	114, 118, 122	127	154	82
4	114, 118, 122, 125	137	154	89
5	114, 115, 118, 122, 125	144	154	94
6	114, 115, 118, 122, 124, 125	147	154	95